

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 05-067137
(43)Date of publication of application : 19.03.1993

(51) Int. Cl. G06F 15/38

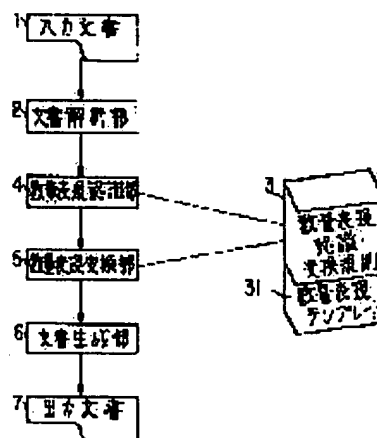
(21)Application number : 03-229893 (71)Applicant : FUJITSU LTD
(22)Date of filing : 10.09.1991 (72)Inventor : NISHINO FUMITO
IBUKI JUN
NAKAMURA NAOTO
ERITSUKU BUI TSUSAA

(54) DOCUMENT QUANTITY EXPRESSION CONVERTER

(57) Abstract:

PURPOSE: To convert quantity expression into appropriate one by searching a document, and using a table designated in advance.

CONSTITUTION: This converter is provided with a quantity expression template 31 consisting of a word 32 including the quantity expression, and a conversion function 34 which converts the quantity expression and the quantity expression to be converted can be fetched from a matching character string when the matching of the quantity expression template 31 with the word 32 including the quantity expression for the character string of a supplied document is taken when it is performed, and it can be converted by the conversion function 34, and it can be replaced with the quantity expression in the document.



32	33	34	31
平成n年	m(YEAR)	f(1988+n)	
長さn	m(LENGTH)	f(256*n)	
昭和n年	m(YEAR)	f(n-1925)	平成n年(昭和n年) 平成n年(昭和n年) 平成n年(昭和n年)

LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's

decision of rejection or application
converted registration]

[Date of final disposal for
application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's
decision of rejection]

[Date of requesting appeal against
examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998, 2003 Japan Patent Office

F-02RL0044

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)特許出願公開番号

特開平5-67137

(43)公開日 平成5年(1993)3月19日

(51)Int.Cl.⁵

G 0 6 F 15/38

識別記号

庁内整理番号

M 9194-5L

F I

技術表示箇所

審査請求 未請求 請求項の数3(全14頁)

(21)出願番号 特願平3-229893

(22)出願日 平成3年(1991)9月10日

(71)出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中1015番地

(72)発明者 西野 文人

神奈川県川崎市中原区上小田中1015番地

富士通株式会社内

(72)発明者 伊吹 潤

神奈川県川崎市中原区上小田中1015番地

富士通株式会社内

(72)発明者 中村 直人

神奈川県川崎市中原区上小田中1015番地

富士通株式会社内

(74)代理人 弁理士 岡田 守弘

最終頁に続く

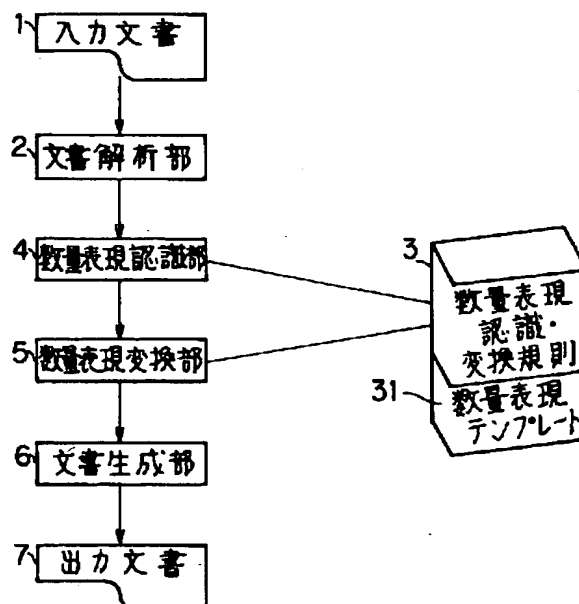
(54)【発明の名称】 文書内数量表現変換装置

(57)【要約】

【目的】 本発明は、文書中の数量表現を変換する文書内数量表現変換装置に関し、文書内をサーチし、予め指定したテーブルを用いて数量表現を適切な数量表現に変換することを目的とする。

【構成】 数量表現を含む単語32、および数量表現を変換する変換関数34からなる数量表現テンプレート31を備え、与えられた文書中の文字列について数量表現テンプレート31の、数量表現を含む単語32とマッチングを行ってマッチしたときに、マッチした文字列から変換すべき数量表現を取り出し、これを変換関数34によって変換し、文書中の数量表現と置き換えるように構成する。

本発明の原理構成図



【特許請求の範囲】

【請求項1】 文書中の数量表現を変換する文書内数量表現変換装置において、

数量表現を含む単語(32)、および数量表現を変換する変換関数(34)からなる数量表現テンプレート(31)を備え、

与えられた文書中の文字列について上記数量表現テンプレート(31)の、数量表現を含む単語(32)とマッチングを行ってマッチしたときに、マッチした文字列から変換すべき数量表現を取り出し、これを上記変換関数(34)によって変換し、文書中の数量表現と置き換えるように構成したことを特徴とする文書内数量表現変換装置。

【請求項2】 文書中の数量表現を変換する文書内数量表現変換装置において、

数量表現を含む単語(32)、および数量表現を変換する変換関数(34)と区間、下方表現、上方表現からなる数量表現テンプレート(31)を備え、

与えられた文書中の文字列について上記数量表現テンプレート(31)の、数量表現を含む単語(32)とマッチングを行ってマッチしたときに、マッチした文字列から変換すべき数量表現を取り出し、これを上記変換関数(33)によって変換した数量表現が上記区間内のときにこの数量表現、上方のときに上方表現した数量表現、あるいは下方のときは上記下方表現した数量表現と置き換えるように構成したことを特徴とする文書内数量表現変換装置。

【請求項3】 文書中の数量表現を変換する文書内数量表現変換装置において、

数量表現を含む単語(32)、数量表現の意味情報(33)、および数量表現を変換する変換関数(34)からなる数量表現テンプレート(31)を備え、

与えられた文書中の文字列について上記数量表現テンプレート(31)の、数量表現を含む単語(32)とマッチングを行ってマッチしたときに、マッチした文字列から変換すべき数量表現を取り出して中間表現で表現し、この中間表現を上記意味情報(33)が同一の目標言語の中間表現に変換し、置き換えるように構成したことを特徴とする文書内数量表現変換装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、文書中の数量表現を変換する文書内数量表現変換装置であって、与えられた文書中の数量表現を、指定された数量表現に変換する文書内数量表現変換装置に関するものである。

【0002】文書中には、“3尺”や“1013mb”のような単位付き数量、“昭和31年”のような年号などの種々な数量化表現が使用されている。これら数量化表現の一部は著者の文化や趣味、専門分野に依存し、これらの文書の文化圏の異なる人間が読むとその数量表現

の表わしている意味を正確に把握できないことがある。この意味の表現するところを正確に把握したり、伝えたりするためには数量表現を変換する必要がある。この数量表現の変換を自動的に行うことが望まれている。

【0003】

【従来の技術】従来から電卓のようなもので単位系の数値を変換する機能を持つものが存在する。これは、人間が文書中の数量表現を探して、数値を入力して変換するものであり、文書中の数量表現を一括して変換するものではない。

【0004】また、テキストエディタや機械翻訳のような文書中の文字列や単語を指定された文字列や単語に置き換えるというものが存在する。これらは単に指定された文字列や単語を変換するのであって、その意味であるところの数量を扱い、自動的に変換するものではない。

【0005】

【発明が解決しようとする課題】上述した電卓のようなものは、文書内に数量表現が現れる毎に、手動でその数量表現を変換する必要があり、手間がかかるという問題がある。また、テキストエディタや機械翻訳のようなものでは、文字列や単語を指定した文字列や単語に変換することはできるが、意味である数量の変換が行えないという問題がある。また、数量表現を機械翻訳などの内部に組み込んだ装置では、利用者が柔軟に新たな数量表現のための規則を定義することができないし、その他のアプリケーションでは容易に使用することができないという問題がある。

【0006】本発明は、文書内をサーチし、予め指定したテーブルを用いて数量表現を適切な数量表現に変換することを目的としている。

【0007】

【課題を解決するための手段】図1は、本発明の原理構成図を示す。図1において、数量表現テンプレート31は、文書中に現れる、数量表現を含む単語32、数量表現の意味情報33、数量表現を変換する変換関数34と区間、下方表現、上方表現からなるテンプレートである。

【0008】数量表現認識部4は、与えられた文書中の数量表現を認識するものである。数量表現変換部5は、取り出された数量表現を新たな数量表現に変換するものである。

【0009】

【作用】本発明は、図1に示すように、数量表現認識部4が与えられた文書中の文字列について数量表現テンプレート31の数量表現を含む単語32とマッチングを行ってマッチしたときに、マッチした文字列から変換すべき数量表現を取り出し、数量表現変換部5がこれを数量表現テンプレート31の変換関数33によって目的とする数量表現に変換し、文書中の数量表現と置き換えるようにしている。

【0010】また、数量表現認識部4が与えられた文書中の文字列について数量表現テンプレート31の数量表現を含む単語32とマッチングを行ってマッチしたときに、マッチした文字列から変換すべき数量表現を取り出し、数量表現変換部5がこれを数量表現テンプレート31の変換関数34によって変換した数量表現が区間内のときにこの数量表現、上方のときに上方表現した数量表現、あるいは下方のときは下方表現した数量表現とし、文書中の数量表現と置き換えるようにしている。

【0011】また、数量表現認識部4が与えられた文書中の文字列について数量表現テンプレート31の数量表現を含む単語32とマッチングを行ってマッチしたときに、マッチした文字列から変換すべき数量表現を取り出し、数量表現変換部5がこれを一旦中間表現で表現し、意味情報33が同一の目標言語の中間表現に変換し、文書中の数量表現と置き換えるようにしている。

【0012】従って、文書内をサーチし、数量表現テンプレート3.1とマッチングを行って数量表現を取り出し、これを変換関数34によって適切な数量表現に変換して文書中の数量表現と置き換えることにより、文書中の数量表現を適切な数量表現に自動的に変換することが可能となる。

【0013】

【実施例】次に、図1から図10を用いて本発明の実施例の構成および動作を順次詳細に説明する。

【0014】図1は、本発明の原理構成図を示す。図1において、入力文書1は、数量表現を変換する対象の入力された文書である。

【0015】文書解析部2は、入力された入力文書1を形態素解析および構文解析などを行うものである。数量表現認識部4は、数量表現認識・変換規則3を参照し、入力文書1の数量表現を認識して取り出すものである。

【0016】数量表現変換部5は、数量表現認識部4によって入力文書1から取り出された数量表現について、数量表現認識・変換規則3の変換規則（例えば変換関数34や区間、下方表現、上方表現など）を用いて適切な数量表現に変換するものである。

【0017】文書生成部6は、数量表現変換部5によって変換された適切な数量表現を、元の入力文書1の数量表現と置き換えて文書として整え、出力文書7を生成するものである。

【0018】以下図2から図10を用いて図1の構成の実施例について順次詳細に説明する。図2は、本発明の数量表現の検索処理フローチャートを示す。

【0019】図2において、S1は、文を読み込む。読み込む文がなければ（S1のNO、EOF）、終了する（END）。S2は、読み込む文字列中のi番目の文字を指すインデックスiを0と初期化する。

【0020】S3は、インデックスiが指す文字がなければ（iが文字列の長さL以上ならば）、これまで選択

した単語列を結果として、S1に戻る。S4は、文字列中のi番目からの文字列とマッチする単語および数量表現テンプレート31を単語辞書から検索する。これは、例えば図3に示す数量表現テンプレート31を予め登録しておき、入力文書の文字列について、数量表現を含む単語32とマッチするものを検索、例えば“平成*n年”（ここでは*nは任意の数字列）をマッチする単語として検索して見つける。

【0021】S5は、適当な単語を選択し、インデックスiを進める。そして、S3に戻る。ここで、適切な単語を選択する方法として、最長一致法（マッチした最長の文字列を単語と選択する手法）、文節数最小法、コスト最小法などのいずれかを採用し、最適な単語を選択する。

【0022】以上の手順によって入力文書1から見つけた数量表現を含む単語（例えば“平成*n年”）を検索して取り出す。図3は、本発明の数量表現テンプレート例を示す。これは、入力文書1から数量表現を検索して取り出し、新たな数量表現に変換する数量表現テンプレート31の例を示す。

【0023】数量表現テンプレート31は、入力文書1から数量表現を検索して取り出し、新たな数量表現に変換するテンプレートであって、数量表現を含む単語32、数量表現の意味情報33、および新たな数量表現に変換する変換関数34などから構成されるものである。

【0024】単語32は、入力文書1とマッチして数量表現を含む文字列を取り出すためのものである。例えば単語“平成*n年”の“*n”は任意の数字を表わし、これとマッチする入力文書の文字列を数量表現を含む単語として取り出す。

【0025】意味情報33は、数量表現の意味を表わす情報である。例えば単語32“平成*n年”のときは、意味情報33は“m(YEAR)”、即ち数量表現が年を表わす。

【0026】変換関数34は、新たな数量表現に変換する変換関数である。例えば単語32“平成*n年”のとき変換関数34“f(1988+*n)”、即ち1988年に平成*n年の*nを加算した結果が西暦となる。

【0027】以上のように、数量表現テンプレート31として、数量表現を含む単語32、数量表現の意味情報33、および新たな数量表現に変換する変換関数34などを予めテンプレートとして登録しておくことにより、入力文書1から図2のフローチャートに従って数量表現を取り出し、これを後述する図4のフローチャートに従って新たな数量表現に変換する。

【0028】図4は、本発明の数量表現変換処理フローチャートを示す。これは、図2のフローチャートによって入力文書1から取り出した数量表現を含む単語32について、新たな数量表現に変換するときの処理フローチャートである。

【0029】図4において、S11は、形態素解析リストからポインタで指される次の単語を取り出す。S12は、終りか否かを判別する。終りのときは終了する(END)。

【0030】S13は、その単語が数量表現の単語か否かを判定し、もし数量表現の単語でなければS11に戻る。これは、形態素解析されたリストから順次取り出した単語について、図3の数量表現テンプレート31の数量表現を含む単語32とマッチングを行い、マッチした単語を数量表現の単語と判定する。

【0031】S14は、数量表現の単語32について、数量表現テンプレート31の意味情報33を取り出し、その単語がどのような種類の単語かを認識する。S15は、S14で取り出した意味情報33をキーとして、変換すべき数量表現を取り出す。

【0032】S16は、変換関数による新しい数量を計算する。これは、元の数量表現の変換関数 $f(x)$ と、変換対象の変換関数 $g(x)$ とから変換後の新たな数量を計算する。この計算は、下式(1)によって計算する。

【0033】

【数1】

$$g^{-1}(f(n)) \text{-----} (1)$$

ここで $g^{-1}(n)$ は $g(n)$ の逆関数

【0034】S17は、元の形態素解析リスト中の数量表現を新しい数量表現の単語で置き換える。以上の処理によって、例えば元の形態素解析リストから取り出した数量表現を含む単語“平成3年”が、図3の単語32

“平成*n年”とマッチし、ここでは変換関数 $f(1988+n)$ を適用して $1988+3=1991$ と数量を求め、新しい数量表現を含む単語“1991年”で置き換える。ここで、置き換えるのではなく、元の数量表現と新たな数量表現とを併記、例えば“平成3年(1991年)”としてもよい。また、“平成*n年”から直接に指定された数量表現の数量が図3の数量表現テンプレート31に登録されていないときは、上述した式

(1)に示すように、元の変換関数 $f(x)$ と、変換対象の変換関数 $g(x)$ をもとに新たな数量を求める。

【0035】図5は、本発明の区間指定の数量表現変換処理フローチャートを示す。これは、図3の数量表現テンプレート31の変換関数34に、変換関数の他に区間情報を予め登録しておき、これを用いて区間に分けて数量表現を変換するものである。例えば数量表現を含む単語31が“*n年”のとき、変換関数34“ $f(*n-1925)$ ”、これに区間b(昭和1、64)、下方表現i(大正)、上方表現s(平成)を登録し、これを用い、西暦を新たな昭和に変換した後、昭和、大正、あるいは平成に変換する。以下説明する。

【0036】図5において、S21は、新しく変換した数量表現の数値が、指定された区間内に収まるか判定する。S22は、収まるならば、その数量表現を利用し、新たな数量表現の単語にして終了する。例えば西暦を昭和に変換した後の数値*Kが1から64の範囲に収まるならば、昭和*K年とする。

【0037】S23は、区間より小さいならば、下方表現を利用することとし、S21、S22によって、新たな数量表現の単語にし終了する。例えば西暦を昭和に変換した後の数値*Kが1から64の範囲よりも小さいならば、S21、S22によって大正*M年とする。

【0038】S24は、区間より大きいならば、上方表現を利用することとし、S21、S22によって、新たな数量表現の単語にし終了する。例えば西暦を昭和に変換した後の数値*Kが1から64の範囲よりも大きいならば、S21、S22によって平成*M年とする。

【0039】以上によって、数量表現テンプレート31の変換関数34に区間、下方表現、上方表現などを予め登録しておくことにより、例えば西暦を昭和、大正、平成などに正しく容易に変換することが可能となる。

【0040】図6は、本発明の中間表現変換による数量表現変換処理フローチャートを示す。これは、入力文書1の数量表現を含む単語を一旦中間表現に変換し、それから目的とする数量表現に変換するものである。

【0041】図6において、S31は、入力された原文の入力文書の数量表現を含む単語について、解析し、中間表現で表現する。例えば図8で後述する原文“二尺”について、この原文“二尺”の言語(日本語)の概念に基づいて図8の(1)の中間表現で表現する(詳細は後述する)。

【0042】S32は、原文の単語の言語の概念に基づいて表現した中間表現の各概念に対して、その概念が数量概念ならば目標となる言語の文の数量概念に変換を行う。例えば図8で(1)の原文の数量概念の中間表現を、図8の(2)、(3)に示すように目標とする言語の数量概念の中間表現に順次変換する。

【0043】S33は、目標言語で正しく表現された中間表現を使用して目標言語の文(単語)を生成する。そして、これを出力する。以上によって、入力文書1から取り出した数量表現を含む原文の文(単語)について、原文の言語の数量概念の中間表現で一旦表現し、次にこれを目標言語の数量概念の中間表現で表現し、これを変換して正しい数量概念の中間表現にした後、目標言語の数量概念の文(単語)を生成することが可能となる。

【0044】図7は、本発明の概念検査による数量概念変換処理フローチャートを示す。これは、入力文書1の数量表現を含む単語を一旦中間表現に変換し、それから目的とする言語の同一の概念をもつかをチェックして検査し、適合するものを目標言語の数量表現の文として生成するものである。

【0045】図7において、S41は、入力された原文の入力文書の数量表現を含む単語について、解析し、中間表現で表現する。例えば図8で後述する原文“二尺”について、この原文“二尺”の言語（日本語）の概念に基づいて（1）の中間表現で表現する（詳細は後述する）。

【0046】S42は、原文の言語の概念に基づいて表現した中間表現に対して、全ての目標言語の概念についてチェックし終わったら、S46に進む。S43は、中間表現から概念を1つ取り出し、取り出した概念が目標言語の概念として存在していたら、S42に戻り、存在していなかったらS44に進む。

【0047】S44は、目標とする言語の概念で、その数量を表現できる概念を概念変換テーブルで探す。この概念変換テーブルには、予め言語に存在する数量概念の対応関係を登録しておく（例えば日本語の数量表現（長さ）“尺”と英語の数量表現（長さ）“centimeter”）S45は、元の中間表現の概念と数値を、目標言語の中間表現の概念と数値に変換する。例えば日本語の数量表現（長さ）“尺”と数値を英語の数量表現（長さ）“centimeter”と数値に変換する。そして、S42に戻る。

【0048】S46は、概念変換した後の中間表現が、目標言語の中間表現として規則正しいか検査する。規則正しいならば、S49に進む。規則正しくなければ、S47で中間表現を調べて中間表現規則の違反理由を見つけ、S48で中間表現の変換を行って正しくし、S46に戻る。

【0049】S49は、目標言語で正しく表現した中間表現を使用し、目標言語の数量表現を含む文（単語）を生成する。そして、出力する。以上の処理によって、日本語の数量表現を含む単語“二尺”を目標言語である英語の数量表現を含む単語“60.6 centimeters”に変換するときの処理を、図8を用いて具体的に説明する。

【0050】図8は、本発明の中間表現例（その1）を示す。図8において、（1）は、日本語の数量表現を含む単語“二尺”を日本語（原文）の中間表現で表わしたものである（図7のS41）。

【0051】ここで、“尺”を表わす概念C#shakuは英語（目標言語）には存在しないので（図7のS43のNO）、英語の同じ長さ概念であるc#centimeterという概念に変換し（図7のS44、S45）、英語の中間表現の（2）を生成する。

【0052】（2）は、（1）の日本語の概念の中間表現から、英語の概念の中間表現に変換したものである。ここで、この（2）の中間表現は、英語の概念を使用して表現されているが、英語の中間表現としては正しくない（図7のS46のNO）ので、数値を計算して中間表現を変換する（図7のS47、S48）。（3）を経て（4）の英語の正しい中間表現に変換する。そして、（4）の英語の中間表現から数量表現を含む文“60.6 ce

ntimeters”を生成する。

【0053】図9は、本発明の中間表現例（その2）を示す。図9において、（1）は、日本語の数量表現を含む単語“昭和63年”を日本語（原文）の中間表現で表わしたものである（図7のS41）。

【0054】ここで、“昭和63年”を表わす概念C#showaは英語（目標言語）には存在しないので（図7のS43のNO）、英語の同じ概念であるc#a.dという概念に変換し（図7のS44、S45）、英語の中間表現の（2）を生成する。

【0055】（2）は、（1）の日本語の概念の中間表現から、英語の概念の中間表現に変換したものである。ここで、この（2）の中間表現は、英語の概念を使用して表現されているが、英語の中間表現としては正しくない（図7のS46のNO）ので、数値を計算して中間表現を変換して（3）にする（図7のS47、S48）。そして、（3）の英語の中間表現から数量表現を含む文“theyear 1988”を生成する。

【0056】図10は、本発明の文生成における数量表現変換処理フローチャートを示す。これは、例えば英語の25 centsのような表現を、quarterに変換するものである。

【0057】S51は、原文を解析して中間表現で表現する。S52は、S51で生成した中間表現から概念を取り出す。S53は、S52で取り出した概念が数量表現の概念か否かを検査する。

【0058】S54は、S53の検査で数量表現の概念でなく通常概念であったとき、通常訳語選択を行い、S57で目的言語の文を生成する。S55は、S53の検査で数量表現の概念であったとき、それと同種の意味の数量表現の概念から指定された数量表現の訳語を取り出す。例えば“25 cents”や“quarter”のうちから“quarter”の訳語を取り出す。

【0059】S56は、生成する数量にマッチするように、数量を計算する。そして、目標言語の文を生成する。

【0060】

【発明の効果】以上説明したように、本発明によれば、文書内をサーチし、数量表現テンプレート31とマッチングを行って数量表現を取り出し、これを変換関数34によって適切な数量表現に変換して文書中の数量表現と置き換える構成を採用しているため、文書中の数量表現を適切な数量表現に自動的に変換することができる。これにより、文書内の数量表現を一括して適切な数量表現に変換することが可能となり、機械翻訳システムや文書推敲支援システムなどに利用することができる。

【図面の簡単な説明】

【図1】本発明の原理構成図である。

【図2】本発明の数量表現の検索処理フローチャートである。

【図3】本発明の数量表現テンプレート例である。

【図4】本発明の数量表現変換処理フローチャートである。

【図5】本発明の区間指定の数量表現変換処理フローチャートである。

【図6】本発明の中間表現変換による数量表現変換処理フローチャートである。

【図7】本発明の概念検査による数量表現変換処理フローチャートである。

【図8】本発明の中間表現例（その1）である。

【図9】本発明の中間表現例（その2）である。

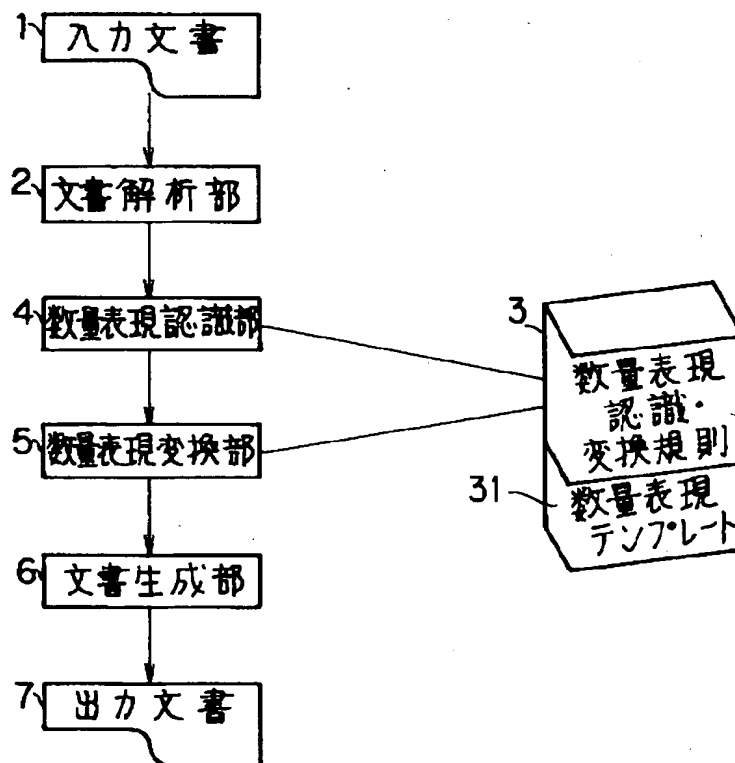
【図10】本発明の文生成における数量表現変換処理フローチャートである。

【符号の説明】

- 1：入力文書
- 2：文書解析部
- 3：数量表現認識・変換規則
- 31：数量表現テンプレート
- 32：数量表現を含む単語
- 33：意味情報
- 34：変換関数
- 4：数量表現認識部
- 5：数量表現変換部
- 6：文書生成部
- 7：出力文書

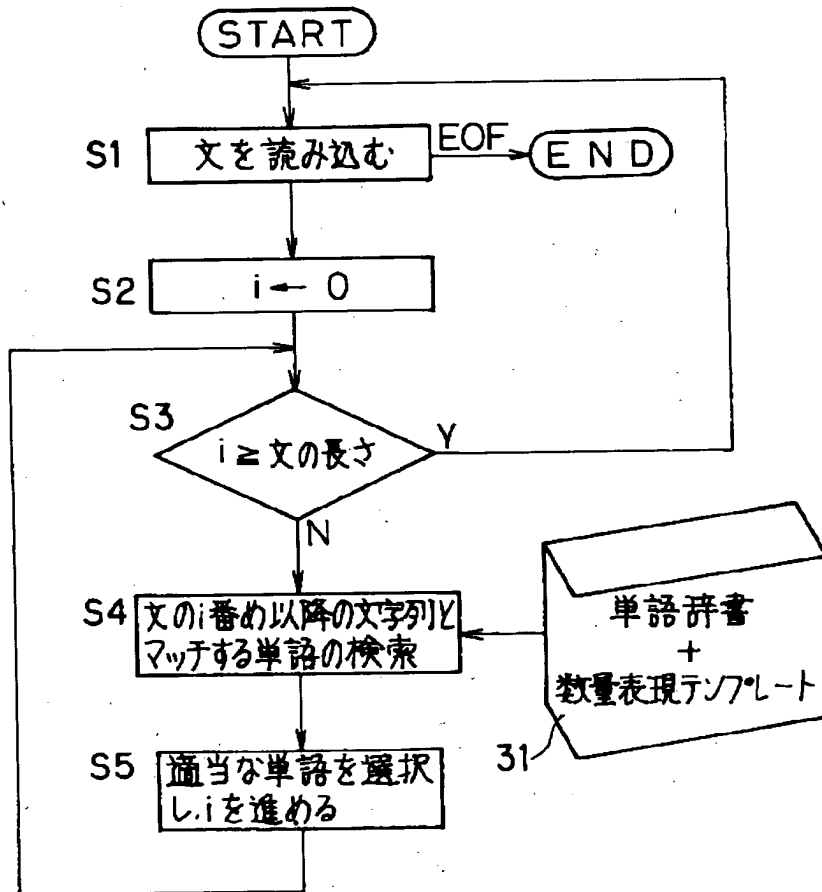
【図1】

本発明の原理構成図



【図2】

本発明の数量表現の検索処理フローチャート



【図3】

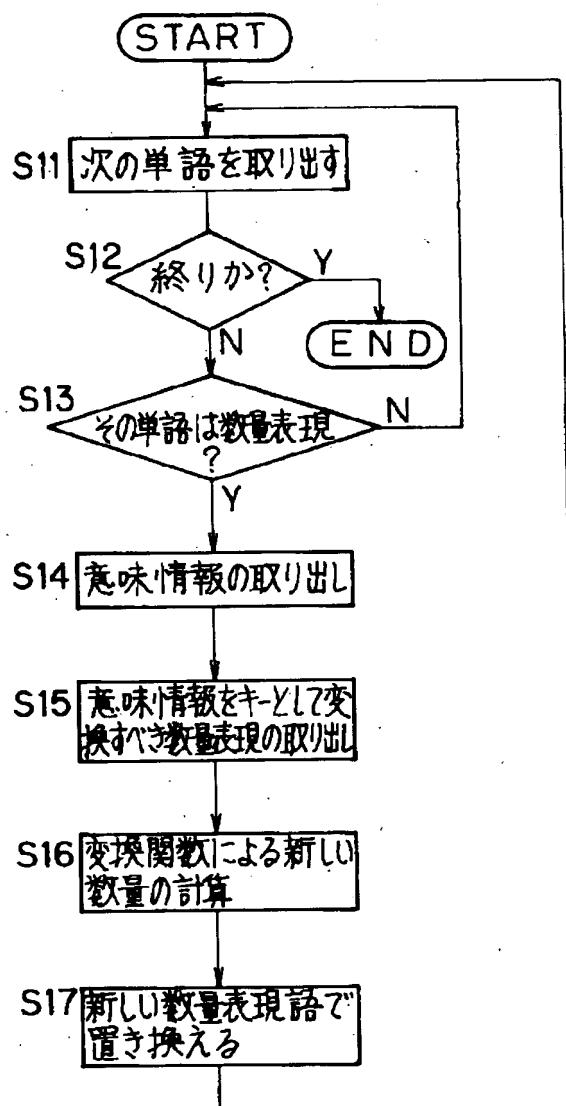
本発明の数量表現テンプレート例

31

32 単語	33 意味情報	34 変換関数	
平成*n年	m(YEAR)	f(1988+*n)	
*nインチ	m(LENGTH)	f(2.54x*n)	
*n年	m(YEAR)	f(*n-1925)	区間b(昭和1.64) 下方表現i(大正) 上方表現s(平成)

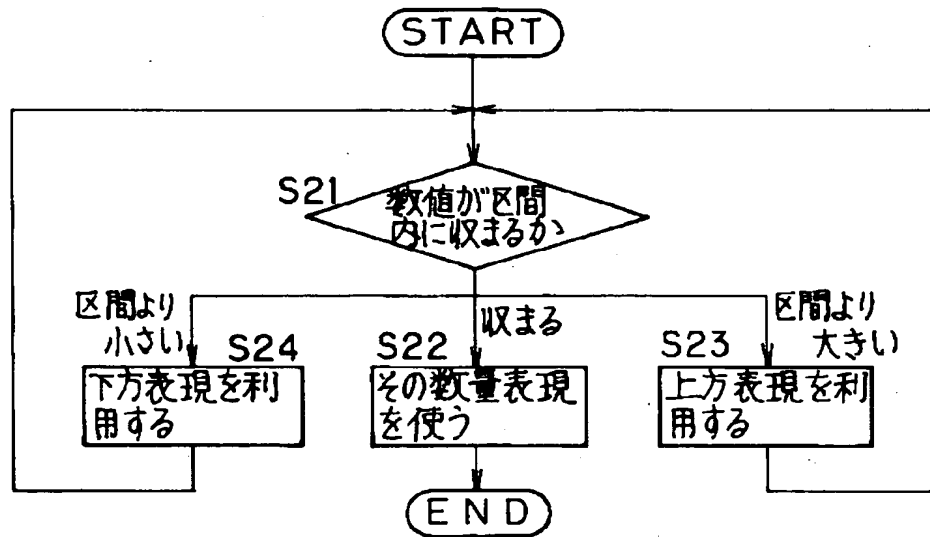
【図4】

本発明の数量表現変換処理フローチャート

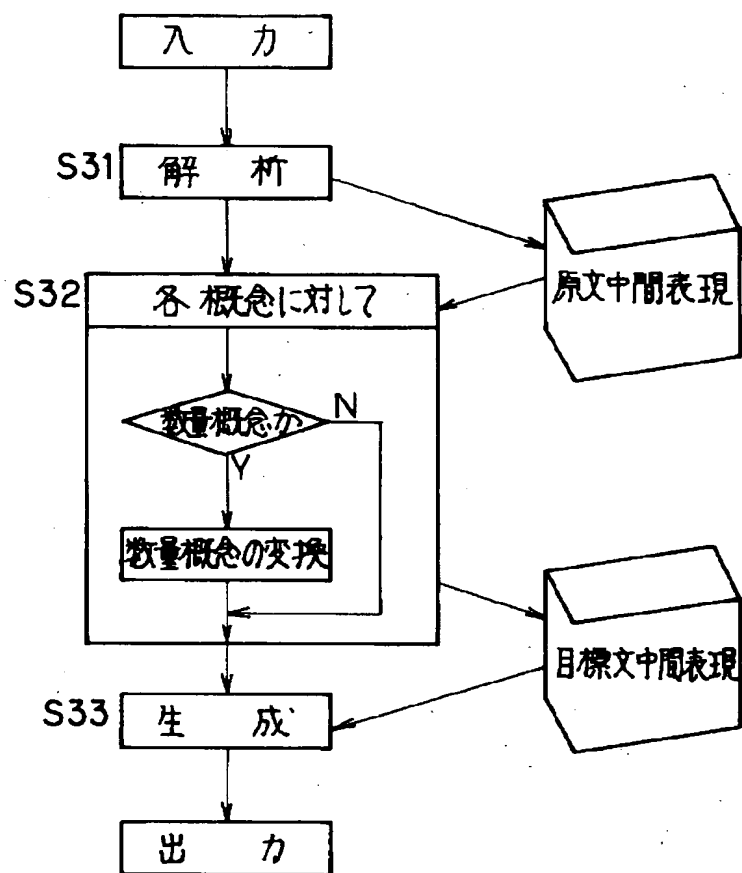


【図5】

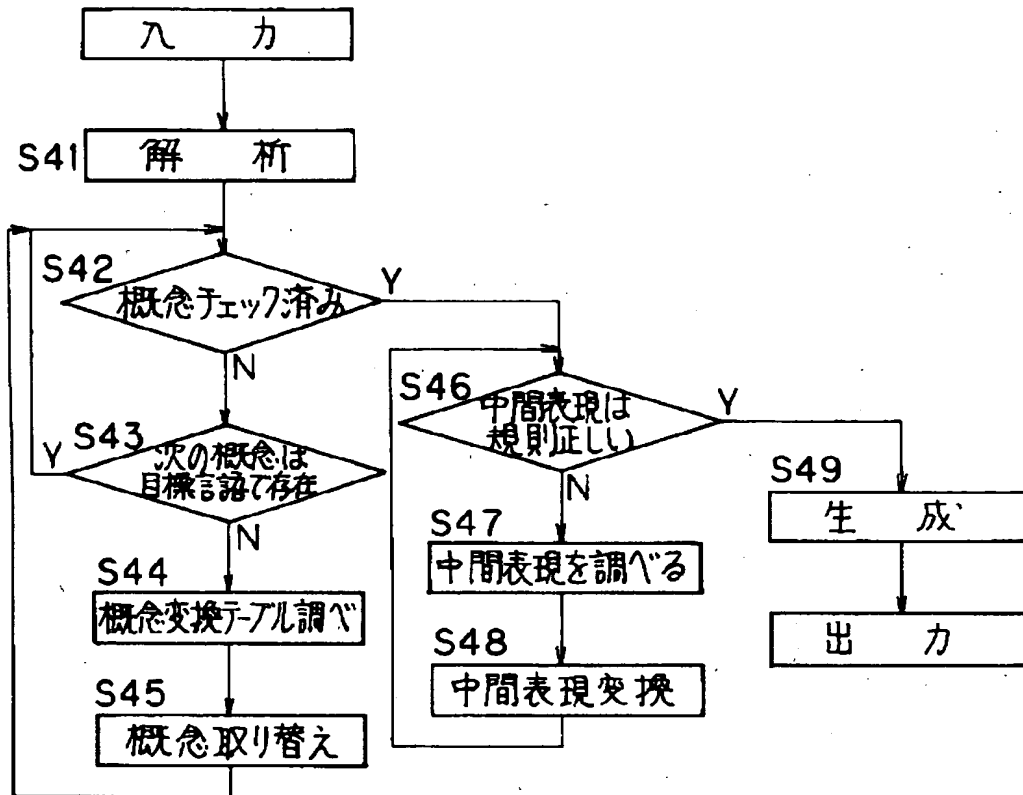
本発明の区間指定の
数量表現変換処理フローチャート



本発明の中間表現変換による
数量表現変換処理フローチャート

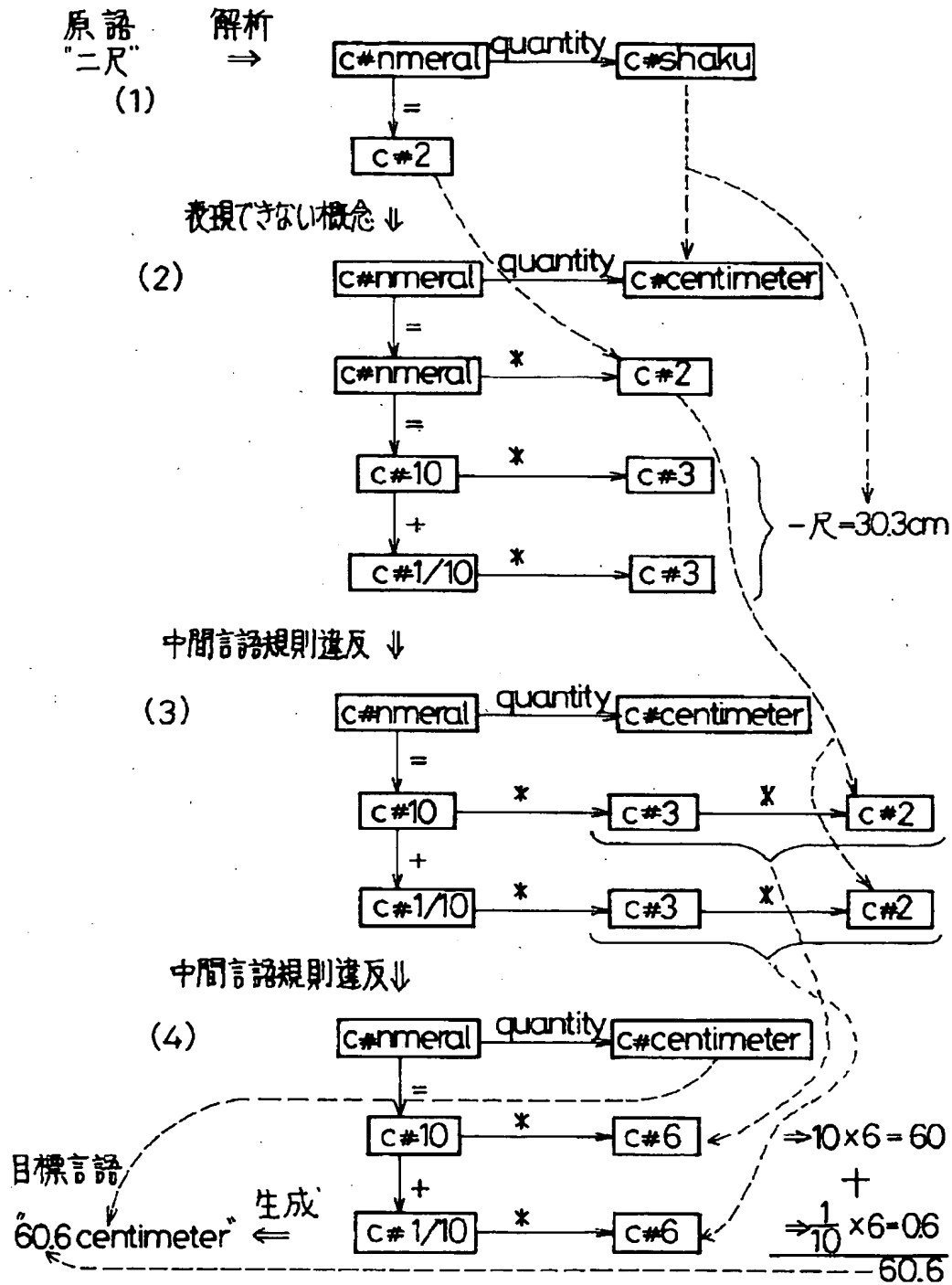


【図7】

本発明の概念検査による
数量概念変換処理フローチャート

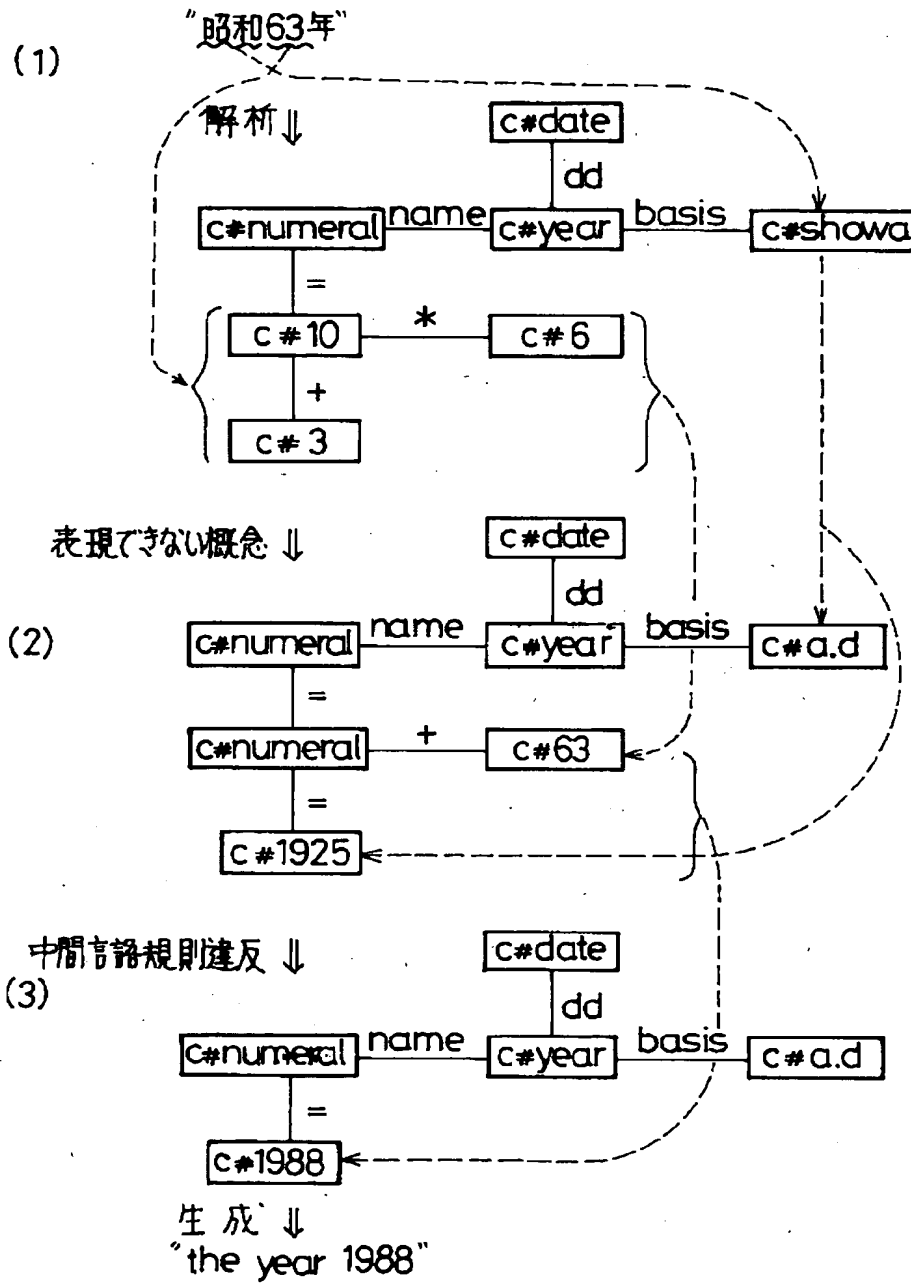
【図8】

本発明の中間表現例(その1)

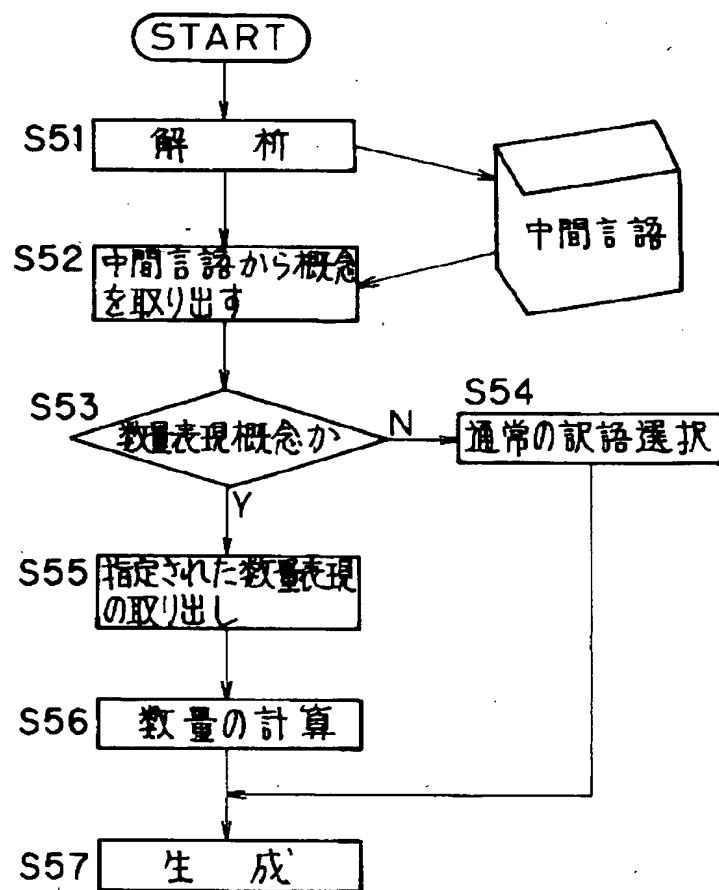


【図9】

本発明の中間表現例(その2)



【図10】

本発明の文生成における数量表現変換
処理フローチャート

フロントページの続き

(72)発明者 エリック ヴイツサー
神奈川県川崎市中原区上小田中1015番地
富士通株式会社内